Thus, the same number of long-range constraints provides relatively less structural information for β-proteins. As a result, the demands on the force field with a given (small) number of constraints are greater. While frequently the proper fold can be identified by choosing the lowest energy final conformation, as happened in the studies reported here, this may not always be the case. Indeed, when the magnitude of the energy fluctuations is larger than the observed energy difference for the final states, a different protocol for the selection of the proper fold is necessary. Such a protocol is described in the next section.

Structure Refinement and Selection of Native Folds

As described above, the lowest-energy final structures from simulated annealing, representing the putative proper fold and the closest competing alternative topology, were subjected to isothermal Monte Carlo runs using the same force field and sets of constraints. The results of these stage 3 runs are summarized in Table V, below.

TABLE V. Compilation of the Results of the Isothermal Simulations

Name	Fold	Average cRMSD	cRMSD (A) Cα-fit	Average energy	$(S)^{1/2}$ (Å)
6pti(9/S1)	NAT[a]	3.69 (0.21)	4.28 (0.29)[c]	-323.4	10.6
41 res	MI[b]	Not observed			
18-56					
6pti(9/S2)	NAT	3.67 (0.22)	3.60 (0.11)	-326.1	10.6
	MI	Not observed			
6pti(9/S3)	NAT	4.41 (0.12)	4.23 (0.07)	-325.0	9.9
	MI	8.01 (0.21)		-301.0	10.6
6pti(9/S4)	NAT	4.01 (0.29)	4.05 (0.22)	-327.6	10.6
	MI	8.11 (0.34)		-309.8	9.6
6pti(9/S5)	NAT	4.06 (0.24)	4.27 (0.14)	-349.7	10.8
	MI	8.34 (0.23)		-318.4	10.2
1gb1(8)	NAT	3.11 (0.13)	3.39 (0.14)	-582.9	10.9
56 res	MI	8.61 (0.13)		-567.2	11.5
1ctf(10)	NAT	3.48 (0.25)	3.21 (0.08)	-699.9	11.2
68 res	MI	8.68 (0.16)		-656.9	11.7
1pcy(46)	NAT	3.44 (0.11)	3.80 (0.08)	-856.5	12.7

99 res	MI	11.34 (0.08)		-796.9	12.7
1 pcy(25)	NAT	4.87 (0.12)	4.88 (0.04)	-952.5	13.1
	MI	Not observed			
1pcy(15)	NAT	5.27 (0.06)	5.70 (0.16)	-891.7	13.0
	MI	7.70 (0.08)		-841.8	12.9
2trx(30)	NAT	3.63 (0.15)	3.11 (0.14)	-1013	13.0
108 res	MI	Not observed			
2trx(16)	NAT	3.43 (0.12)	3.52 (0.06)	-1082	13.3
	MI	11.88 (0.11)		-888	13.2
4fab(27)	NAT	4.77 (0.06)	4.42 (0.07)	-1040	13.8
111 res	MI	11.49 (0.13)		-1011	13.8
4fab(16)	NAT	5.53 (0.08)	5.92 (0.10)	-1137	14.1
	MI	12.76 (0.09)		-1033	13.8
3fxn(35)	NAT	3.91 (0.12)	4.06 (0.08)	-1514	14.3
138 res	MI	12.94 (2.33)		-1311	14.3
3fxn(20)	NAT	4.44 (0.22)	4.12 (0.14)	-1401	14.3
	MI	Not observed			
1mba(20)	NAT	4.44 (0.23)	4.34 (0.05)	-1698	15.0
146 res	MI	Not observed			
Atim(62)	NAT	5.19 (0.10)	5.08 (0.11)	-2423	17.4
247 res	MI	Not observed			
Atim(50)	NAT	5.77 (0.06)	5.96 (0.04)	-2483	17.7
	MI	Not observed			
Atim(36)	NAT	6.66 (0.09)	6.74 (0.15)	-2622	17.9
	MI	9.97 (0.05)		-2549	17.9

[a] Native structure.

[b] Misfolded (generally the topological mirror image fold) structure.

[c] The number in parentheses is the standard deviation of the coordinate root-mean-square distance in Angstroms between the crystallographic and predicted α-carbon traces; *see also* the legend for Table IV, above.

All simulations were done at T = 1. The average cRMSD from native and the average energy are computed from 200 snapshots of the Monte Carlo trajectory. In all cases, the proper fold can be identified based on the average conformational energy. Thus, a combination of fast-simulated annealing and long isothermal runs allows the dependable selection of the proper fold. Indeed, during rapid assembly via Monte Carlo-simulated annealing, a fine-tuning of structural details is not always achieved. In long isothermal runs, the misfolded (topological mirror image conformations) states could always be detected as those of higher average

conformational energy. For the case 6pti where five different sets of constraints were examined, the lowest energy misfolded structure has a higher conformational energy than the highest energy proper fold, regardless of the set of constraints. On average, the accuracy of the predicted native fold improved slightly during the isothermal runs and ranges between 3 and 5 Å cRMSD (for the estimated positions of the alpha-carbons), except for the Atim barrel where it was about 6 Å. By way of illustration, Figures 8 and 9 present a representative conformation (generated using the MOLMOL[42] procedure) of 3fxn and 4fab obtained from the isothermal refinement runs (employs 20 and 16 constraints, respectively) with a cRMSD of 4.4 Å and 5.5 Å, respectively.

Increasing the number of long-range constraints, on average, leads to some increase in the compactness of the obtained structures, as assessed by their average root-mean-square radius of gyration. There is no obvious systematic difference between the dimensions of the native and misfolded states. Since in both cases the majority of constraints are always satisfied, the difference in conformational energy arises from the underlying force field that has a reasonable level of specificity for nativelike structures. Unfortunately, the non-constraint contributions to the potential are not sufficiently specific to fold the protein (except for a few small proteins) without the assistance of the constraints. On the other hand, within the limit of N/7 constraints, if the constraints are used alone without the remainder of the potential, the resulting structures are essentially random. Thus, it is the synergism of the constraints with the underlying contributions to the potential that permits the folding of these proteins. For some of the test proteins, good folds could be obtained with a smaller than N/7 constraints (*e.g.*, 4 constraints for protein G). The value of N/7 is a conservative estimate of a safe lower bond for all proteins. This number is smaller than required by related methods.[4-6]

Next, the side-chain-based lattice models serve as targets for building models with two united atoms per residue, *e.g.*, as in the CAPLUS model. Table VI, below, displays the cRMSD data for such reconstructed main chains.